Supplementary Table 1. Details of *M. tuberulosis* isolates sequenced for this study.

Enclosed electronic excel file

Supplementary Table 2. Details of publicly available *M. tuberulosis* genomes included in this study.

Enclosed electronic excel file

Supplementary Table 3. Homoplasic non-synonymous SNPs identified as occurring on a lineage-defining branch and also arising independently within other lineages. The number of branches on which each SNP was identified outside the lineage is shown, and the number of such branches that have multiple descendant tips (indicating onward transmission of the SNP) is shown (no. transmitted). HCMC refers to the 1,635 isolates from HCMC; Elsewhere refers to the 3,085 additional isolates from published studies; ^ indicates nsSNP in T cell epitope region. Panel (a) is reproduced from Table 2 in the main text for ease of comparison.

a) Beijing Lineage 2.2 defining mutations				
	HCMC		Elsewhere	
Mutation	no. branches outside L2.2	no. transmitted	no. branches outside L2.2	no. transmitted
EsxW-T2A	9	4	10	6
Rv3081-F220L	2	1	7	3
GidB-E92D	1	1	0	0

b) Lineage 1 defining mutations				
	HCMC		Elsewhere	
Mutation	no. branches outside L1	no. transmitted	no. branches outside L1	no. transmitted
EsxO-G22A	1	1	0	0
Rv0107c -1247SP	1	1	0	0
MmpL3-302GS	2	0	0	0
Rv0457c-119W*	1	0	0	0
Rv0610c-305Q	1	1	0	0
VapC31-M90V	1	0	0	0
Rv0988-V57I	1	0	0	0
Rv1640c-I949V	1	0	0	0
Rv2813-I76V	1	0	0	0
^ Rv2994-W68*	1	1	0	0

c) Lineage 4 defining mutations				
	HCMC		Elsewhere	
Mutation	no. branches outside L4	no. transmitted	no. branches outside L4	no. transmitted
EsxI-L20Q	1	0	0	0
PrpD-G3D	1	0	0	0
Rv1248c-M764V	1	1	0	0
Rv2825c-W162S	1	0	0	0
Rv3898c-Q111*	2	0	0	0

Supplementary Table 4. *esxW* primers used for PCR and capillary sequencing.

Primer name	Primer sequence
esxW F1	5' TCG CGG TCA ACA CAT CAG AA 3'
esxW R1	5' CTG CCC CTC GAT ACG GAT TC 3'

Supplementary Table 5. Repetitive regions of the *M. tuberculosis* genome that were excluded from SNP analysis.

Enclosed electronic excel file